The Hidden Influence of Robots: How Robot Interaction Strategies Shape Human-Human Interaction and Perception in Creative Ideation

Hristian Semerdzhiev[1]

Supervisor(s): Catharine Oertel[1]**, Ruben Weijers**[1]
[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 21, 2026

Name of the student: Hristian Semerdzhiev
Final project course: CSE3000 Research Project
Thesis committee: Catharine Oertel, Ruben Weijers, Guohao Lan

Abstract—**Social robots deployed in collaborative groups reshape interaction between human participants, yet how different verbal strategies produce distinct shaping effects remains untested. This study compared assertive and supportive robot strategies across 20 dyads in a creative ideation task with a Pepper robot. Transcripts were coded using Mercer's talk taxonomy and post-session interviews were analysed thematically. The supportive strategy produced a significantly higher frequency of exploratory discourse than the assertive strategy. The interaction strategy had no significant effect on participation balance, disputational talk, cumulative talk, group cohesion, or ingroup identification. Four unintended cross-condition shaping effects emerged, including attention redistribution, partner solidarity, speech formalisation, and emotional suppression. Condition-specific effects were also observed, including creative suppression and expert deference. These results show that a robot's verbal register shapes the quality of human collaboration, and that robot presence alone restructures interaction beyond any programmed design intent.**

1 Introduction

Creative ideation rarely happens alone. Whether student teams design solutions to open problems or engineers brainstorm product concepts, the most productive ideas emerge through collaborative conversation, where contributions are challenged, refined, and built upon across many turns (Rosenberg-Kima et al., 2020). The quality of that conversation matters. Discourse in which participants justify reasoning, question assumptions, and elaborate on each other's ideas is linked to stronger collaborative outcomes than mere agreement or assertion (Mercer, 2004). Social robots are entering precisely these collaborative spaces, deployed as brainstorming facilitators (Geerts et al., 2021), co-creation partners (de Rooij et al., 2024), and group learning assistants (Alves-Oliveira et al., 2019).

Placing a robot in a group, however, does not simply add a new contributor. It reshapes the dynamics between the human participants. Gillet et al. (2024) define Interaction-Shaping Robotics (ISR) as the study of robots that influence the behaviors and attitudes exchanged between two or more other agents. A growing body of work confirms that robot behavior propagates into the human-human layer of a group, reshaping who speaks, how evenly turns are distributed, and how participants relate to one another (Sebo et al., 2018; Traeger et al., 2020; Gillet et al., 2021; Lin et al., 2026). Each of these studies, however, isolates a single robot behavior and tests it in one condition, making it impossible to determine how different strategies compare in their shaping effects. Without a cross-condition design, it remains unknown whether a given shaping effect reflects the particular strategy a robot uses or simply the fact of having a robot present at all. Nor has prior work systematically distinguished between effects robots are designed to produce and those that emerge unintentionally from their presence in the group.

This gap is especially noticeable in creative ideation. Geerts et al. (2021) showed that substituting a robot for a human facilitator does not reliably change brainstorming productivity. They suggest that the specific strategy the robot uses matters more than its mere presence. In creative tasks, what is at stake is not only who speaks, but how constructively participants talk. Comparing strategies on both of these dimensions, within the same task and setting, requires a design that prior work has not yet attempted.

The central question of this thesis is:

> ***How do different robot interaction strategies shape human-human interaction and perception in creative ideation?***

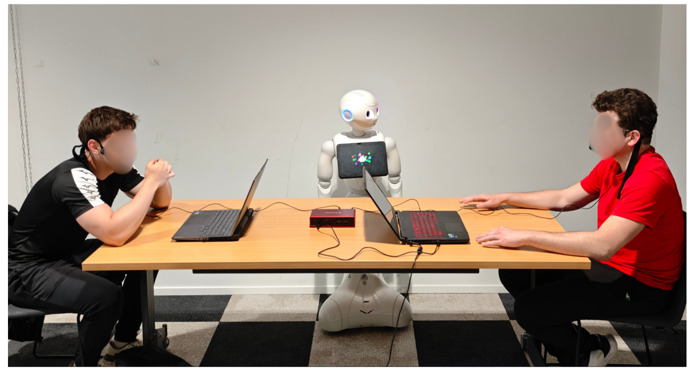


Figure 1: Experimental setup showing two participants interacting with Pepper during the brainstorming task. Participant faces have been anonymized for privacy.

We pursue this question through four sub-questions:

SQ1 How does the inequality of word counts among human partners vary across different robot interaction strategies?

SQ2 How does the proportion of human utterances (disputational, cumulative and exploratory) vary across robot interaction strategies?

SQ3 How does the robot's interaction strategy influence participants' reported group cohesion and ingroup identification with their human partner?

SQ4 What shaping effects emerge across conditions, and do they differ between strategies?

We address these questions through a between-subjects experiment in which a Pepper robot was deployed in one of two verbal interaction strategies, assertive or supportive, across 20 dyads engaged in a collaborative ideation task. No silent-robot or no-robot control was included. The study is designed to compare the two strategies against each other rather than against a baseline of unaided conversation. Cross-condition analysis of session transcripts and post-session interviews examined the structure and quality of human-human conversation as well as participants' subjective accounts of the interaction.

Our results show that the robot strategy significantly altered the verbal quality of dialogue by eliciting a higher frequency of exploratory talk in the supportive condition ($F(1, 18) = 6.16$, $p = .023$). In contrast, the interaction strategy had no measurable effect on participation balance, disputational talk, cumulative talk, self-reported group cohesion, or ingroup identification. Qualitative interview analysis identified four cross-condition shaping effects that emerged in both experimental groups as well as additional conversational dynamics unique to each strategy.

The main contributions of this work are: (1) empirical cross-condition comparison of assertive and supportive robot strategies in a creative group ideation task; (2) a mixed-methods methodology for evaluating interaction-shaping profiles, operationalising creative discourse quality quantitatively through Mercer's talk taxonomy while capturing intended and emergent shaping effects qualitatively via thematic interview analysis; and (3) concrete design implications for robots deployed in collaborative creative settings, grounded in both quantitative interaction measures and qualitative participant reports.

This thesis is structured as follows. Chapter 2 reviews related work on social robots in creative facilitation, formalises the framework of interaction-shaping robotics, and introduces collaborative discourse quality as a distinct dimension that prior ISR work has not yet measured. Chapter 3 presents the research methodology, including the experimental design and measurement instruments. Chapter 4 reports the results of the study, which are then interpreted and discussed in Chapter 5 in relation to existing literature, including limitations and directions for future work. Chapter 6 concludes the thesis. Finally, Chapter 7 addresses responsible research considerations.

2 Related Work

This thesis builds on three strands of prior work. We first review what is known about robots in creative and collaborative tasks, establishing why the specific strategy of the robot matters for ideation quality. We then examine how robots shape human-human interaction dynamics, situating the present work within the ISR framework. Finally, we show that the quality of collaborative talk is a distinct and consequential dimension of group interaction that prior ISR work has not yet measured, and we connect it to the questions the present study addresses.

2.1 Robot Facilitation of Creative and Collaborative Tasks

A central question in robotic group facilitation is whether robots can meaningfully improve collaborative and creative group processes. Recent work suggests that the answer depends not simply on the presence of a robot, but on how it participates in the interaction. A recent scoping review identified 108 publications targeting changes in human-human interaction through robotic facilitation, spanning 85 distinct application targets (Weisswange et al., 2026). Despite this breadth of research, evidence for the effectiveness of robotic facilitation remains mixed.

For example, Geerts et al. (2021) compared brainstorming sessions facilitated by either a robot or a human ($N = 54$) and found no significant difference in idea productivity. Simply replacing a human facilitator with a robot was therefore insufficient to improve creative performance. This finding suggests that the specific interaction strategy adopted by the robot may be more important than its mere presence.

Supporting this view, de Rooij et al. (2024) showed that a robot facilitator's mood expressions induced corresponding mood contagion in participants that shaped both collaboration quality and creative output ($N = 110$), with positive expressions producing the strongest improvements. Where they varied the robot's affective register, Rosenberg-Kima et al. (2020) varied the robot's functional role and found that robot facilitation of small-group activities in higher education was associated with more positive attitudes toward the activity and perceived advantages in time management over a human facilitator.

In an ideation-specific setting, Pham et al. (2024) showed that a gesture-based robot promoted more consensus-building behaviour in group sessions. Their post-session interviews further revealed that participants felt socio-emotionally supported without consciously attributing this to the robot's actions, pointing to the importance of capturing emergent effects that go beyond the robot's stated design intent.

Together, these studies establish how a robot participates shapes collaborative creative processes. However, none directly compares multiple robot interaction strategies under the same task and conditions. Whether a directive, assertive strategy produces different collaborative discourse than a warm, supportive one, and whether those differences extend beyond what the robot's design targets, remains an open question that the present work addresses.

2.2 Interaction-Shaping Robotics

The field of ISR, focuses on how robots can influence interpersonal dynamics. To better understand these effects, it is helpful to examine the structural factors that define a robot's influence and the specific sensory channels it uses.

The Structural Factors of Robotic Influence

Gillet et al. (2024) organise ISR according to five factors that determine how a robot shapes interaction between other agents. The first two concern the robot's position in the group: its role, which ranges from guiding facilitator to peer group member, and its type of communication, either verbal or nonverbal. The third factor is the form of influence, which can be explicit, meaning the robot directly prompts a change in the interaction, or implicit, meaning it acts indirectly without naming its intent. The fourth factor is the robot-shaping outcome, which can be behavioural, such as a change in speaking time or gaze behaviour, or cognitive, such as a change in attitudes, feelings, or trust. The fifth factor is the timeline, distinguishing effects that unfold immediately from those that persist after the robot's shaping behaviour has concluded.

These five factors help organise the evidence reviewed in this section. The studies below differ primarily along two dimensions: the type of communication the robot uses (verbal or nonverbal) and the form of its influence (explicit or implicit). Prior work has predominantly targeted behavioural outcomes such as participation balance and speaking time (Gillet et al., 2021; Komura et al., 2024), or cognitive outcomes such as trust and group cohesion (Birmingham et al., 2020; Weisswange et al., 2026). The present study targets a different outcome: the quality and distribution of collaborative discourse types. It examines how varying the robot's verbal strategy shifts that outcome, a combination of factors that prior ISR work has not tested.

Verbal and Nonverbal Shaping Strategies

Research on verbal shaping strategies provides some of the clearest evidence of downstream effects on human-human interaction. Sebo et al. (2018) showed that a robot's expressions of vulnerability produced ripple effects on team trust, with human teammates becoming more likely to console one another, explain failures, and laugh together during moments of tension. Traeger et al. (2020) subsequently extended this line of work by demonstrating that groups whose robot made vulnerable statements also spoke substantially more with each other and distributed turns more evenly than groups with a neutral or silent robot. Both findings confirm that a robot's verbal register shapes not just how people respond to the robot itself, but how they engage with one another.

Nonverbal strategies follow a parallel logic. Gillet et al. (2021) showed that adaptive robot gaze shifted participation toward less active speakers in a language game, and Tennent et al. (2019) extended this principle to a peripheral robotic object, demonstrating that a device without speech improved group engagement and problem-solving through implicit nonverbal cues alone. Komura et al. (2024) further showed that calibrating nonverbal signals to social hierarchy can reduce dominant speakers' talking time without significantly decreasing their reported satisfaction.

Our work differs from these studies in both the robot behavior and the interaction outcomes examined. Rather than using nonverbal cues to regulate participation or speaking time, we contrast supportive and assertive verbal interaction strategies. This allows us to investigate how robot behavior shapes the quality of human-human dialogue, particularly the emergence of exploratory, cumulative, and disputational talk.

2.3 The Quality of Collaborative Talk

Prior work has demonstrated that robots can influence the quantity of human participation, which is typically measured through speaking time, turn counts, and participation balance (Tennent et al., 2019; Asano et al., 2024). However, these metrics say comparatively little about the quality of what participants say. In creative ideation, this distinction matters greatly. A group where both partners contribute equal word counts but only agree with each other produces a fundamentally different outcome from a group where they challenge, justify, and build on each other's proposals. The question of which robot strategy produces more constructive discourse cannot be answered by participation metrics alone.

To address this limitation, we draw on the foundational work of Mercer et al. (1999), who argues that the quality of collaborative talk is central to group performance and learning. They demonstrate that groups whose talk is characterised by reasoning and joint evaluation develop stronger learning and collaborative results than groups whose talk consists of uncritical agreement or unresolved disagreement.

To capture these qualitative differences systematically, Mercer (2004) provides a theoretically grounded framework known as Sociocultural Discourse Analysis, which distinguishes three distinct types of collaborative talk. First, disputational talk consists of disagreement without constructive reasoning. Second, cumulative talk consists of uncritical agreement and elaboration. Third, exploratory talk consists of critical, co-constructive exchange in which participants justify ideas, evaluate alternatives, and refine contributions together.

This framework positions the quality of collaborative exchange as an independent dimension of group interaction. It makes creative ideation a particularly consequential setting for studying robot interaction strategies because what matters is not only that participants speak, but that they speak in ways that generate, refine, and challenge ideas. Prior work in human-robot interaction has paired structured interaction coding with post-session interviews to surface discourse-level effects in group ideation settings (Pham et al., 2024). The present work follows that precedent by applying the Mercer taxonomy as a cross-condition measure of discourse quality, while directing the interview analysis toward how participants experienced each other's contributions rather than focusing only on the robot.

3 Methodology

This study adopted a between-subjects experimental design to compare how two distinct robot interaction strategies shape human-human interaction in a creative ideation context. Two conditions were constructed around the same task and platform, with the robot's interaction strategy as the only systematically varied factor. This design allows any differences in conversational dynamics to be attributed to the strategy rather than to the task, the setting, or the robot hardware.

Data were collected through session transcripts and post-session questionnaires and interviews. Thus, yielding both quantitative measures of interaction structure and qualitative accounts of how participants experienced the conversation.

An overview of the study design, procedure, and data streams is shown in Figure 2.

3.1 Participants

A total of 40 participants (33 male, 7 female) participated in the study. The mean age was 22.8 years (SD = 2.4). Participants comprised 26 Bachelor's students, 13 Master's students, and 1 PhD student. They were recruited using a convenience sampling method through social networks and on-campus recruitment, including approaching students on the spot at TU Delft's campus. Inclusion criteria required that participants were fluent in English and were at least 18 years old. Pairs were formed randomly, resulting in 20 dyads across 2 experimental conditions.

3.2 Experimental Setup

Task

Two participants and a Pepper[1] robot were seated together at a table, arranged as shown in Figure 1. A laptop displaying the task sheet (see Appendix B) was placed in front of participants at the start of the session.

The task asked them to brainstorm ideas for improving campus life, with suggested topics: student well-being, social interaction, food and services, events and activities, and safety and comfort. These topics were available as prompts if the conversation stalled. The goal was to generate as many meaningful and creative ideas as possible.

Sessions lasted approximately 12 minutes. Audio was captured using individual headset microphones worn by each participant. A camera recorded the session for reference.

Robot Interaction Role

The Pepper robot participated as an active contributor in the brainstorming conversation rather than as a passive facilitator. Its responses were generated at runtime using the *Phi-3.5-mini-3.8b-instruct* large language model, prompted according to the assigned interaction strategy for that condition (see Figure 3).

The robot operated proactively: it could intervene without being directly addressed. Interventions were triggered by one of four conditions: (1) a participant addressed Pepper by name; (2) a participant uttered a predefined struggle cue (e.g., *"we're stuck"*, *"I don't know"*); (3) seven or more seconds of silence elapsed; or (4) a detection algorithm identified that the four most recent participant turns had introduced fewer than six new content words relative to the preceding eight turns, indicating the discussion was circling rather than progressing.

10 dyads were assigned to the assertive condition and 10 to the supportive condition. The two strategies differed in both verbal framing and vocal delivery parameters.

In the assertive condition, the robot used a direct and commanding tone, pushed participants toward new directions, and did not hesitate to challenge or redirect ideas. Example utterances included phrases such as *"You need to think about..."* or *"Consider this..."*. Speech was delivered at a faster rate (120% of the NAOqi default speed) and at a moderate volume.

In the supportive condition, the robot used a warm and encouraging tone, validated participants' contributions before extending them, and offered gentle prompts when participants struggled. Example utterances included *"That's a great point, and it makes me think of..."* or *"I see where you're coming from, and it could also be interesting to consider..."*. Speech was delivered more slowly (80% speed) and at a softer volume to match the empathetic register of the condition.

3.3 Procedure

Upon arriving at the experiment site, participants were asked to read and sign an informed consent form (see Appendix A). Participants who did not consent were not permitted to continue and were thanked for their time.

Following consent, participants were introduced to the session format without being told the robot's specific role, in order to limit demand effects. The group conversation with Pepper lasted approximately 12 minutes.

[1] Pepper is a humanoid social robot developed by SoftBank Robotics. See https://us.softbankrobotics.com/pepper

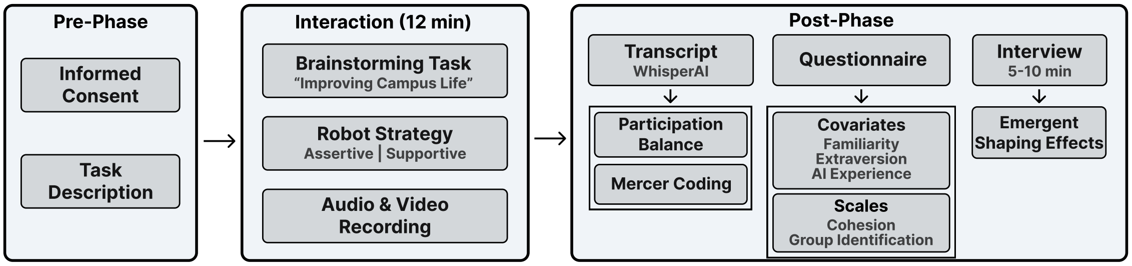


Figure 2: Study design overview. Each brainstorming session yielded three data streams: speaker-labelled audio transcripts used to compute participation balance and Mercer talk-type distributions (SQ1 and SQ2), post-session questionnaires addressing group cohesion and ingroup identification (SQ3) and semi-structured interviews capturing shaping effects on human-human interaction (SQ4).

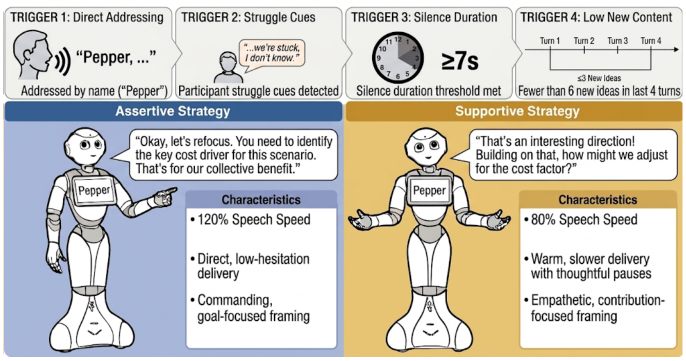


Figure 3: System overview of Pepper's proactive intervention architecture. The top row shows the four intervention triggers (direct addressing, struggle cues, $\geq$7s silence, and low content progression), while the bottom panels contrast the behavioral characteristics and speech parameters of the assertive and supportive strategies.

After the interaction, each participant individually completed a post-session questionnaire (see Appendix C). The questionnaire collected demographic information (age, gender, and educational background) and measures of familiarity between participants, prior robot experience, and extraversion to account for potential confounding influences on group interaction. The primary dependent variables, group cohesion and ingroup identification, were measured using validated scales from Henry et al. (1999) and Chin et al. (1999).

Subsequently, both participants were invited to take part in a short semi-structured interview lasting approximately 5-10 minutes. Interview questions focused on how participants perceived the robot and how they experienced their human partner during the conversation. For example, the participants were asked: "*Did having the robot there change how open you felt sharing your own experiences?*". The interview guide, including follow-up probes, is provided in Appendix D.

Participants were thanked upon completion. The full procedure lasted approximately half an hour.

3.4 Mixed-Methods Analysis

Data were analysed using a mixed-methods approach, combining quantitative transcript coding and questionnaire data with qualitative thematic analysis of interview material. These three streams address complementary sub-questions. The transcript analysis captures measurable differences in interaction structure across robot-strategy conditions (SQ1 and SQ2). The questionnaire data address the relational effects of strategy on group cohesion and ingroup identification (SQ3). The thematic analysis surfaces participants' subjective accounts of emergent changes in how they interacted with their human partner (SQ4).

Transcript Coding

Speaker-labelled transcripts of both the group conversations and post-task interviews were generated using a locally hosted instance of OpenAI Whisper[2]. Audio recordings were transcribed with the Whisper medium model, and speaker diarization was used to assign utterances to individual speakers.

Utterances were coded according to Mercer (2004) taxonomy of talk, which distinguishes between disputational, cumulative, and exploratory talk. Unlike word-count-based productivity measures or generic content analysis, this taxonomy captures the quality of collaborative reasoning rather than its volume, making it particularly suited to a study whose central question concerns how participants think together rather than how much they say.

Inter-rater reliability was assessed by having a second coder independently code the same 16% sample of utterances. Cohen's κ was computed to be .76. A minimum threshold of $\kappa \geq .70$ was required before proceeding with full coding. A complete coding manual, including decision rules and worked examples, is provided in Appendix E.

Participation balance was quantified using the adjusted Gini coefficient (Deltas, 2003) as operationalised by Asano et al. (2024), computed over per-speaker word counts within each session. For a session with n speakers and word counts $x_1, x_2, \ldots, x_n$, the adjusted Gini coefficient is defined as:

$$G = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n \sum_{i=1}^{n} x_i} \tag{1}$$

where the $\frac{n}{n-1}$ factor corrects for the downward bias of the standard Gini in small groups. A value of $G = 0$ reflects perfectly equal participation. Values approaching 1 indicate increasing concentration of talk in one speaker.

The adjusted Gini coefficient was preferred over simpler alternatives such as standard deviation of word counts or raw turn counts because it produces a normalized inequality score that is independent of session length and total words spoken. Asano et al. (2024) validate its use for participation balance in small collaborative groups specifically because of this property.

Group Cohesion and Group Identification

Group cohesion was operationalised using the six-item Perceived Cohesion Scale (Chin et al., 1999), comprising two three-item subscales measuring group belonging and group morale. Ingroup identification was measured using the twelve-item Tripartite Model scale (Henry et al., 1999), with four items each assessing affective, behavioural, and cognitive identification and two items being reverse-scored per scale. All items used a seven-point Likert scale (1 = *Strongly Disagree*, 7 = *Strongly Agree*), and subscale scores were computed as item means. Internal consistency was assessed using Cronbach's α and McDonald's

[2]OpenAI Whisper. Available at: https://openai.com/index/whisper/

ω prior to analysis. Both scales were preferred over single-item self-report measures or custom instruments because their multi-item structure provides stronger internal consistency and allows subscale-level interpretation.

Shaping Effects

Post-session interview audio was transcribed using the same Whisper pipeline as the session transcripts. The resulting transcripts were analysed using thematic analysis following the six-phase procedure of Braun and Clarke (2006): familiarisation, initial code generation, theme search, theme review, theme definition and naming, and report production. Thematic analysis was selected over deductive coding approaches because it permits themes to emerge inductively from participant accounts rather than being constrained to categories anticipated by the robot's design. This property is essential when the research goal is to surface emergent rather than intended effects.

The unit of analysis was the participant's attributed account of a change in how they talked or related to their human partner. Passages addressing only the robot's perceived behaviour were excluded unless the participant explicitly linked them to a shift in human-human interaction. Following initial thematic identification, effects were classified along two dimensions: whether they were *intended* (aligned with the condition's design goals) or *unintended* (not programmed into the robot's behaviour), and whether they were *cross-condition* (emerging in both strategies) or *condition-specific*. The interview guide is provided in Appendix D.

4 Results

The four sub-questions are addressed in turn below. Together, the results show a clear dissociation in the effects of robot interaction strategy. Strategy influenced how participants communicated, particularly in the quality of collaborative talk, while having no measurable impact on how evenly they distributed speech or how they evaluated their group experience.

The interview analysis provides a complementary perspective on these quantitative findings. Across conditions, four consistent shaping effects on human-human interaction emerged, alongside additional effects that were specific to each interaction strategy.

4.1 Participation Balance (SQ1)

We evaluated participation balance to determine how evenly the conversational floor was shared between the human partners across the experimental conditions. Before running our primary statistical models, we conducted preliminary covariate screening to check if participant extraversion, prior robot experience, or pre-existing familiarity influenced conversational symmetry. The linear regression analyses for this screening phase are presented in Tables 3 and 4 of Appendix F.1. Because none of these baseline participant characteristics were significantly correlated with the participation metrics, we removed the covariates from the primary analysis to ensure a parsimonious model.

We then evaluated the distribution of the adjusted Gini index to verify the assumptions required for parametric testing. A Shapiro-Wilk test revealed that the normality assumption was significantly violated (see Appendix G, Table 22. To address this violation appropriately, we analyzed the participation balance using a non-parametric Mann-Whitney U test instead of a standard ANOVA. This analysis showed that the robot interaction strategy did not have a significant effect on the adjusted Gini index ($U = 37.00$, $p = .353$). Conversational inequality remained statistically equivalent across both supportive and assertive robot conditions.

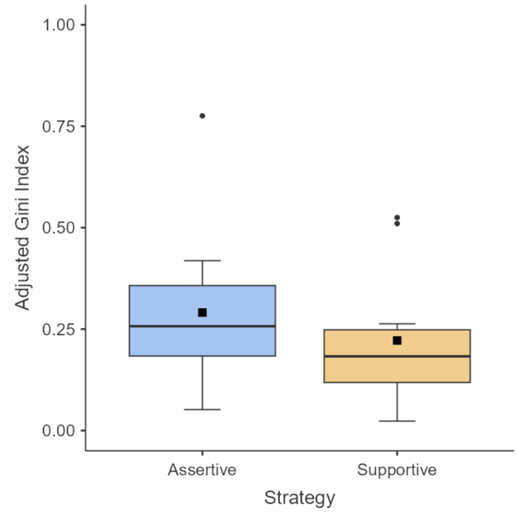


Figure 4: Adjusted Gini index by robot interaction strategy. Strategy had no significant effect on participation balance ($U = 37.00$, $p = .353$; see Table 1). Each box shows the interquartile range The horizontal line marks the median and the square shows the mean. Whiskers extend to $1.5\times$ IQR. Individual points are outliers.

Table 1: Mann-Whitney U test results for the adjusted Gini index by robot strategy. Strategy did not significantly affect participation balance.

	Mann-Whitney U	p
Strategy	37.00	.353

4.2 Quality of Collaborative Talk (SQ2)

We next examined the quality of collaborative verbal interactions by analyzing the specific categories of discourse defined by the Mercer taxonomy. Similar to our participation analysis, we performed preliminary linear regressions to evaluate the influence of the baseline covariates on disputational, cumulative, and exploratory talk. As documented in Tables 5 through 10 of Appendix F.2, none of the baseline covariates demonstrated a significant relationship with any of the talk types. We therefore excluded the covariates from these models and utilized simple one-way ANOVA test for our primary evaluations because the normality assumptions for these verbal metrics were fully satisfied (see Appendix H, Table 23).

The one-way ANOVA test shown in Table 2 revealed a significant effect of the robot interaction strategy on the frequency of exploratory talk ($F(1, 18) = 6.16$, $p = .023$). Human dyads in the supportive robot condition produced more exploratory talk than dyads in the assertive robot condition. In contrast, no significant effects of the robot strategy were found for disputational or cumulative talk. The differences between conditions for these latter two discourse categories were negligible and did not reach statistical significance.

Table 2: One-way ANOVA results evaluating the effect of robot interaction strategy on Mercer talk types. Strategy significantly affected exploratory talk but not disputational or cumulative talk.

	F (1, 18)	p	η^2p	ω^2
Disputational	1.54	.230	0.079	0.026
Cumulative	0.13	.719	0.007	-0.045
Exploratory	6.16	.023	0.255	0.205

4.3 Group Cohesion and Group Identification (SQ3)

The post-session questionnaires show robot strategy produced no significant effect on any of the five sub-scales measuring

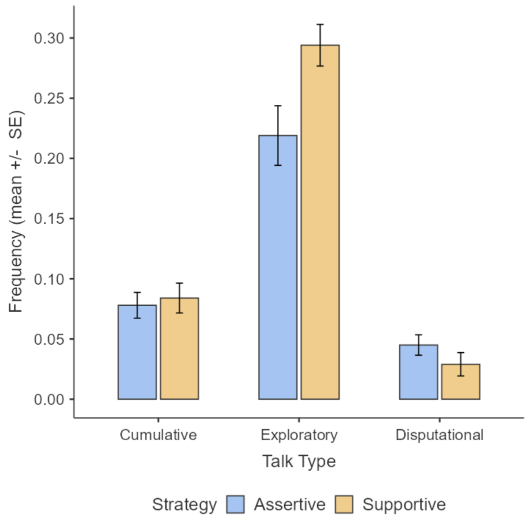


Figure 5: Distribution of Mercer talk types by robot strategy. Strategy was significantly associated with exploratory talk (F(1,18) = 6.16, p = .023, η_p^2 = .255; see Table 2). The supportive condition produced more exploratory and fewer disputational episodes than the assertive one.

cohesion and ingroup identification (full results in Appendices I and K). Regardless of which strategy participants encountered, they reported similar levels of group belonging, group morale, and affective, behavioral, and cognitive identification.

Scale reliability was strong for the cohesion instrument ($\alpha = 0.90$ and $\omega = 0.91$) and acceptable for the ingroup identification instrument overall ($\alpha = 0.71$ and $\omega = 0.75$), with the exception of the behavioral sub-scale, which showed poor internal consistency ($\alpha = 0.41$ and $\omega = 0.53$) and should be interpreted with caution (see Appendices J and L).

Two covariates did reach significance, however, and both point to meaningful patterns. Participants with greater conversational AI experience reported stronger group belonging, independently of which condition they were in (Appendix I, Table 25). A plausible explanation is that participants unfamiliar with AI spent more attentional resources monitoring the robot, leaving fewer resources for their human partner. Dyads who already knew each other beforehand reported stronger cognitive identification with the group (Appendix K, Table 35), suggesting that existing relational history matters more for group feeling than the verbal register the robot used.

4.4 Shaping Effects (SQ4)

The thematic analysis of post-session interviews identified two categories of effects: those that were aligned with each condition's design goals, and those that emerged without being programmed into the robot's behavior. The latter are of particular interest because they reveal how the robot's presence reshaped the human-human layer of the interaction in ways that went beyond what the design anticipated.

Intended Shaping Effects

Creative Steering. In the assertive condition, the robot was designed to push participants toward unexplored ideational territory by challenging and redirecting their contributions. Post-session interviews confirmed that this directive approach expanded creative output beyond what the dyads could generate independently. For instance, Participant 8 noted "I wouldn't have thought about the closed bridges, the closed pathing, if the robot hadn't suggested something. It definitely influenced my ideas in that sense" (P8). This interactive framing also sharpened the internal cognitive processing of the participants because the machine's proposals forced them to look deeper. This was observed when Participant 10 noted that the robot's active presence "lets you think about things you probably didn't recognize in the first place, or think about in the first place" (P10).

Directed Attention. The assertive strategy also achieved the intended effect of directed attention as participants actively oriented their focus toward the proposals made by the machine. This shift in conversational attention was explicitly described by a participant who stated that "once the robot joined, it became the focus of the conversation" (P3).

Ideational Scaffolding. In the supportive condition, both intended effects were confirmed by at least seven of ten dyads. The robot extended the ideational reach beyond what participants perceived they would have generated alone. This was highlighted by participants who remarked that "the robot introduced a whole new level to our ideas" (P32), and "when we were stuck, we would have been stuck for longer. Pepper kind of filled in that role" (P36).

Topical Anchoring. The supportive robot also provided effective topical anchoring by preventing tangential drift and keeping the dyads focused on the core brainstorming task. Participants appreciated this regulating presence, with one individual noting that "it kind of prevented us from going down a tangent" (P39).

Unintended Shaping Effects: Cross-condition

Four unintended effects appeared in both conditions.

Attention Redistribution. Attentional resources shifted toward the robot in both conditions. In the assertive condition, participants noted that "I kept the feeling like, we both just focus more on him" (P19/P20). This pattern was mirrored in the supportive condition, where a participant explained "I was so focused on the robot that this distracted the conversation" (P30).

Partner solidarity. The robot's unpredictability prompted partners to coordinate their behavioral management strategies. This shared challenge created an unscripted form of interpersonal alignment. In the assertive condition, one pair covertly shifted to their native language and noted that "we didn't agree to do this beforehand but we know ourselves well enough" (P1/P2). Similarly, a participant in the supportive condition described how the situation forced them into "a collaborative role in both of us trying actively to get whatever we want from the robot" (P38).

Speech formalisation. Both conditions produced slower, simpler, and more structured speech directed not only at the robot but at the partner. In the assertive condition, this pressure meant that "we had to be more concise with each other" (P17). Participants in the supportive condition experienced a similar shift in delivery, noting that "it was more direct and clear. When we just talk with friends we don't really care how we deliver our messages" (P24).

Emotional suppression. Participants in both conditions withdrew personal and emotional content they would have shared in an unmediated dyadic conversation. An assertive condition participant remarked that "I just on purpose did not go over the personal things" (P19). This tendency to withhold emotional depth was also explicit in the supportive condition, where a participant simply stated that I don't talk about my emotions to the AI" (P38).

Unintended Shaping Effects: Condition-specific

Three unintended effects were exclusive to the assertive condition.

Role renegotiation. Despite its directive design, the assertive robot was progressively recast as a passive oracle. Instead of treating the machine as a conversational leader, they repositioned it as an external reference tool. This was illustrated when a group began treating the robot as "…more like a supporter. We talk about things and we're like, oh, what do you think?" (P10).

Creative suppression. Participants attributed reduced generativity to the robot's topic-steering, counteracting its intended creative influence. They reported that the rigid style restricted their conceptual breadth. This constraint was highlighted by Participant 7, who felt that "the robot kind of took away from our creativity the variety of ideas we would have given" (P7).

One unintended effect was exclusive to the supportive condition. ***Expert deference.*** The warm and validating register unintentionally led human participants to treat the system as an authoritative knowledge source. This perception constrained their independent brainstorming, as detailed by a participant who admitted that "I was thinking that he is the expert in this field because he has all this information" (P40).

5 Discussion

The central question of this thesis was how robot interaction strategies shape human-human interaction and perception in creative group ideation. The results answer this question at three levels, and the answer is different at each level.

Strategy profoundly shaped the quality of how participants talked to each other, while leaving the distribution of the floor and participants' sense of group belonging essentially unchanged. Taken together, these findings suggest that robot strategy does not exert a uniform influence on group interaction, but instead operates through distinct mechanisms that affect three different outcomes to different degrees.

Beyond these three outcome layers, the thematic analysis surfaces something equally important. There were seven ways in which the robot reshaped conversation that neither strategy was designed to produce, pointing to the limits of intentional design in human-robot group contexts.

5.1 Absence of Participation Effects

The most intuitive place to look for a robot's influence on group interaction is who speaks and for how long. Yet, strategy had no measurable effect here, and understanding why reveals an important boundary condition for ISR.

Prior studies that found robot-induced participation effects all operated in groups of three or more, where one member was consistently quieter and the robot could target that imbalance. Tennent et al. (2019) showed that a peripheral robotic object produced more even backchanneling among participants in three-person groups through implicit engagement behaviors alone. Meanwhile, Traeger et al. (2020) found that a vulnerable robot produced more equal talk distributions across three-person teams. In our dyads, that structural target does not exist. Two people in a creative brainstorm are already roughly symmetric in opportunity, leaving neither strategy with a participation gap to close.

The null result on participation balance is therefore a consequence of the dyadic setting. This finding adds a useful boundary condition to the ISR literature: strategy-driven participation effects may require an existing imbalance to act upon. The question of how robot strategies shape talk quality moves to center stage.

5.2 Supportive Strategy Promoted Exploratory Talk

While the robot strategy did not alter participation volume, it strongly influenced the collaborative register of the groups. This stands as the central finding for the discourse quality outcome because exploratory talk serves as the primary catalyst for creative ideation. The supportive strategy effectively fostered this co-constructive communication style. When participants justify ideas, challenge proposals, and build on each other's contributions, they are doing exactly what a creative brainstorming task requires. Mercer et al. (1999) link this form of talk to stronger collaborative outcomes. This study demonstrates that a robot's verbal register directly influences whether a group adopts these productive discourse patterns.

The most likely mechanism is the tone of the interaction the robot established. De Rooij et al. (2024) showed that a robot facilitator's positive mood expressions significantly enhanced group collaboration and co-creation performance, suggesting that the affective register a robot introduces shapes how participants engage with each other. Our result extends this from the affective register to the verbal one. It is specifically the warm, validating tone of the supportive strategy, and not just positive affect in general, that sustains the exploratory mode of discourse.

The assertive robot, by contrast, introduced two effects that together explain the lower exploratory rate. Rhythm disruption broke participants' ideational chains, forcing them to hold ideas in suspension while waiting for the robot to respond, and creative suppression meant that the robot's topic-steering narrowed rather than widened their generative range. Both effects replaced the open-ended exploration of ideas with shorter, more reactive turns directed at managing the robot.

Geerts et al. (2021) already showed that a robot facilitator alone does not increase brainstorming productivity relative to a human facilitator. Our data specify that a robot's effectiveness depends on whether its verbal strategy facilitates exploratory discourse, rather than its mere presence. While the supportive strategy creates these conditions, the assertive strategy undermines them.

5.3 Absence of Cohesion and Identification Effects

The absence of strategy effects on cohesion and ingroup identification completes the dissociation picture. This result is consistent with the I-C-E framework (Abrams and der Pütten, 2020), which describes cohesion and ingroup identification as group processes that develop and consolidate through accumulated shared experience rather than crystallising within a single interaction. A twelve-minute brainstorming session may simply be too short and too unfamiliar a context for strategy differences to register as differences in group feeling.

What is more informative are the two covariates that did reach significance. Participants with greater conversational AI experience reported stronger group belonging, independently of which condition they were in. A plausible explanation connects directly to the attention redistribution effect identified in SQ4. Participants unfamiliar with AI spent more attentional resources monitoring and interpreting an unfamiliar agent, leaving fewer resources for their human partner. This reduced attention to the partner may have in turn reduced the sense of group belonging.

The second covariate, prior partner familiarity, predicted stronger cognitive identification. Dyads who already knew each other reported feeling more like a group, suggesting that existing relational history provides the shared context that a single robot-mediated session cannot build from scratch.

Together, these two covariates indicate that the factors most strongly predicting how participants feel about their group are not what the robot says, but the relational and experiential resources participants bring into the session.

5.4 Emergent Shaping Effects

Perhaps the most revealing part of this study is not what the two strategies differed on, but what they shared.

The discovery of four shared cross-condition effects indicates that the physical presence of a robot shapes human communication independently of its programmed verbal register. This directly addresses the critical empirical gap identified within the interaction-shaping robotics framework by Gillet et al. (2024). It proves that certain conversational modifications are structural rather than strategic.

The simultaneous occurrence of speech formalization and emotional suppression demonstrates that an active robot functions as a prominent social audience. This audience status elevates the perceived formality of the environment. Consequently, human partners adapt by systematically slowing down their speech rate and rationing personal disclosures to protect their perceived conversational competence in front of the agent.

The partner solidarity effect is the most structurally interesting finding. Participants jointly managed the robot as a shared challenge, generating unscripted dyadic coordination that was not part of either robot's design. This contrasts directly with the trends observed by Sebo et al. (2020). In their study of verbal support, the robot effectively replaced peer backchanneling and ultimately reduced ingroup solidarity. Within this project, the limitations and unpredictability of the robot provoked solidarity instead of replacing it. Taken together, these contrasting studies suggest that the trajectory of group solidarity depends entirely on participant perception. Ingroup solidarity may decrease if participants view the robot as a resource that replaces their need to support each other. Conversely, solidarity increases when the group views the machine as a collaborative challenge around which they must align.

Beyond these shared effects, the assertive and supportive robots each produced a condition-specific unintended consequence that offset their own intended creative effect. The assertive robot produced creative suppression, with participants reporting that its topic-steering narrowed rather than widened their generative range. The supportive robot produced expert deference, with participants treating its warm, validating tone as a signal of expertise and constraining their own ideas within the robot's implicit agenda. Both effects echo the pattern identified by Sebo et al. (2020): targeted robot behavior can benefit one dimension of group dynamics while unintentionally degrading another.

5.5 Implications for Robot Design

Taken together, the findings carry a clear message for designers of robots deployed in creative collaboration settings.

A supportive verbal strategy is the better choice. It sustains the exploratory discourse that creative ideation depends on, avoids creative suppression, and does not fracture conversational rhythm. The cost of the supportive strategy is expert deference, but this can be mitigated by designing the robot's contributions as questions rather than assertions, reducing the authority signal while preserving the validating register.

An assertive strategy produces occasional creative steering that participants themselves valued, but this comes bundled with role renegotiation, and creative suppression. This is too high a cost for a creative ideation setting where the humans' own contribution is the primary output.

5.6 Limitations

Three limitations should be considered when interpreting these findings. Participants noted that the robot's speech was delivered too quickly and that it failed to recognise context-specific references during the sessions. These issues are known to reduce perceived competence and trust in social robots (Martin et al., 2020; Koike et al., 2026), and some of the shaping effects identified qualitatively, particularly creative suppression and expert deference, may therefore reflect communicative limitations of the platform as much as the programmed strategy itself.

Beyond speech quality, the robot's responses were solution-oriented. This constrained the depth of joint sense-making available to participants, which matters especially for a study whose central finding concerns the quality of exploratory discourse. More capable language model integration has been shown to substantially improve engagement and interaction quality in similar settings (Smit et al., 2024).

The participant sample also presents a generalisability concern. Recruitment through convenience sampling at TU Delft's campus produced a heavily male-skewed group of 33 male and 7 female participants. This reflects the demographic reality of the university, but limits how broadly the findings can be applied. Collaborative communication styles and responsiveness to a robot's verbal register may vary across more diverse populations. The present sample does not allow these differences to be examined or accounted for.

5.7 Future Work

Two directions follow naturally from this study. The most immediate concerns non-verbal behaviour. Participants across both conditions consistently identified the absence of nodding, facial expression, and responsive gesture as a barrier to perceiving Pepper as a genuine conversational partner. A controlled comparison between a static and a gesture-enriched facilitation condition, holding verbal strategy constant, would isolate this contribution and determine whether the structural shaping effects observed here are amplified or attenuated when the robot reads as more socially present.

The second direction concerns the temporal stability of the effects. Single-session studies cannot capture whether the unintended effects identified here, particularly attention redistribution and expert deference, persist once the novelty of the robot diminishes or dissolve with repeated exposure. Long-term designs involving multiple sessions with the same dyads would allow the shaping profile of each strategy to be traced over time, which matters practically for any setting where robots are intended as sustained rather than one-off collaborators.

6 Conclusion

This thesis set out to answer how different robot interaction strategies shape human-human interaction and perception in creative group ideation. The results deliver a clear answer at three levels. Robot strategy strongly shaped the quality of collaborative discourse. The supportive strategy produced a significantly higher frequency of exploratory talk. Strategy left participation balance and reported group cohesion unchanged, revealing that these dimensions are governed by the relational resources participants bring into the session rather than by the robot's verbal style. Beyond the intended strategy-specific effects, four unintended shaping effects emerged in both conditions, demonstrating that some of the most fundamental changes to human-human interaction are structural rather than strategic.

These results deliver on the three contributions stated in Chapter 1. This work provides an empirical cross-condition comparison of assertive and supportive robot strategies in a creative group ideation task. It introduces a replicable methodology for separating intended from emergent shaping effects, operationalised through Mercer's talk taxonomy as a measure of discourse quality and thematic interview analysis as a window into participants' lived experience of the interaction. Finally, it yields actionable design guidance. A supportive verbal register sustains the exploratory discourse that creative ideation requires, while an assertive strategy introduces disruptions that narrow the human participants' own generative range.

7 Responsible Research

This section highlights the core principles of responsible research that have guided this study, focusing on integrity, reproducibility, replicability, ethical conduct, and transparent acknowledgment of limitations throughout the research process.

7.1 Research Integrity and Participant Protection

As detailed in Section 3.3, this study adhered to ethical protocols regarding informed consent and the right to withdraw. To minimize demand effects and ensure experimental integrity, participants were not informed about the assigned interaction strategy. However, no false information about the study goals or the robot's capabilities was provided.

7.2 Reproducibility and Open Science

This study aligns with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016) by providing the experimental setup, prompt structures, and analysis scripts in GitHub[3]. Task description, questionnaires, interview guides, and the coding manual are further documented in Appendices B, C, D and E.

A key challenge for reproducibility is the balance between the "Accessibility" of data (Wilkinson et al., 2016) and participant privacy. While analysis code is open-source, raw recordings and transcripts are stored on the 4TU repository which is accessible only to the research team. To uphold ethical integrity, these data were pseudonymized during analysis and will be permanently deleted after presenting the work to the thesis committee to protect participant anonymity.

7.3 Use of Artificial Intelligence Tools

Artificial intelligence systems were used in two distinct ways during this thesis. First, generative AI tools were used during the writing process for proofreading, improving clarity, and resolving technical LaTeX issues. Second, the Pepper robot used the *Phi-3.5-mini-3.8b-instruct* model to generate conversational responses during the experimental sessions.

AI-generated outputs were not accepted uncritically. All thesis content, coding decisions, interpretations, and conclusions were reviewed and verified by the researcher. Similarly, the prompts controlling Pepper's behaviour were manually designed and constrained to align with the intended interaction strategy of each experimental condition.

7.4 Limitations for Replicability

Exact replication is limited by the stochastic nature of Large Language Models and the situated social dynamics of group interaction. While prompting structures are provided to encourage replicability, variations in participant personality and interpersonal familiarity may influence outcomes independently of the robot's programmed strategy.

References

Abrams, A. M. and der Pütten, A. M. R.-v. (2020). I-c-e framework: Concepts for group dynamics research in human-robot interaction: Revisiting theory from social psychology on ingroup identification (i), cohesion (c) and entitativity (e). *International Journal of Social Robotics*, 12(6):1213–1229.

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: A field study. *ACM Transactions on Human-Robot Interaction*, 8.

[3]GitHub: https://github.com/hsemerdzhiev/Pepper

Asano, Y., Litman, D., King-Shepard, Q., Maidment, T., Langley, T., Davison, T., Nokes-Malach, T., Kovashka, A., and Walker, E. (2024). What metrics of participation balance predict outcomes of collaborative learning with a robot? *arXiv preprint arXiv:2405.11092*.

Birmingham, C., Hu, Z., Mahajan, K., Reber, E., and Matarić, M. J. (2020). Can i trust you? a user study of robot mediation of a support group. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8019–8026. IEEE.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3:77–101.

Chin, W. W., Salisbury, W. D., Pearson, A. W., and Stollak, M. J. (1999). Perceived cohesion in small groups. *Small Group Research*, 30:751–766.

de Rooij, A., van den Broek, S., Bouw, M., and de Wit, J. (2024). Co-creating with a robot facilitator: Robot expressions cause mood contagion enhancing collaboration, satisfaction, and performance. *International Journal of Social Robotics*, 16:2133–2152.

Deltas, G. (2003). The small-sample bias of the gini coefficient: Results and implications for empirical research. *The Review of Economics and Statistics*, 85(1):226–234.

Geerts, J., de Wit, J., and de Rooij, A. (2021). Brainstorming with a social robot facilitator: Better than human facilitation due to reduced evaluation apprehension? *Frontiers in Robotics and AI*, 8.

Gillet, S., Cumbal, R., Pereira, A., Lopes, J., Engwall, O., and Leite, I. (2021). Robot gaze can mediate participation imbalance in groups with diferent skill levels. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 303–311. IEEE Computer Society.

Gillet, S., Vázquez, M., Andrist, S., Leite, I., and Sebo, S. (2024). Interaction-shaping robotics: Robots that influence interactions between other agents. *ACM Transactions on Human-Robot Interaction*, 13.

Henry, K. B., Arrow, H., and Carini, B. (1999). A tripartite model of group identification: Theory and measurement. *Small group research*, 30(5):558–581.

Koike, A., Okafuji, Y., and Song, S. (2026). Practical insights into designing context-aware robot voice parameters in the wild. In *Proceedings of the 21st ACM/IEEE International Conference on Human-Robot Interaction*, pages 1140–1149.

Komura, K., Ozaki, K., and Yamada, S. (2024). Robot can reduce superior's dominance in group discussions with human social hierarchy. In *HAI 2024 - Proceedings of the 12th International Conference on Human-Agent Interaction*, pages 242–249. Association for Computing Machinery, Inc.

Lin, T. H., Kopelman, Y. R., Busse, M., Sebo, S., and Erel, H. (2026). The impact of a robot's agreement (or disagreement) on human-human interpersonal closeness in a two-person decision-making task. *Computers in Human Behavior*, 174.

Martin, F. A., Malfaz, M., Álvaro Castro-González, Castillo, J. C., and Ángel Salichs, M. (2020). Four-features evaluation of text to speech systems for three social robots. *Electronics*, 9:267.

Mercer, N. (2004). Sociocultural discourse analysis: analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1:137–168.

Mercer, N., Wegerif, R., and Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25(1):95–111.

Pham, T. V., Weisswange, T. H., and Hassenzahl, M. (2024). Embodied mediation in group ideation - a gestural robot can facilitate consensus-building. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, pages 2611–2632, New York, NY, USA. Association for Computing Machinery.

Rosenberg-Kima, R. B., Koren, Y., and Gordon, G. (2020). Robot-supported collaborative learning (rscl): Social robots as teaching assistants for higher education small group facilitation. *Frontiers in Robotics and AI*, 6.

Sebo, S., Dong, L. L., Chang, N., Lewkowicz, M., Schutzman, M., and Scassellati, B. (2020). The influence of robot verbal support on human team members: Encouraging outgroup contributions and suppressing ingroup supportive behavior. *Frontiers in Psychology*, 11:590181.

Sebo, S. S., Traeger, M., Jung, M., and Scassellati, B. (2018). The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 178–186. IEEE Computer Society.

Smit, K., Leewis, S., Almoustafa, H., Yildirim, K., and Uymaz, T. (2024). Enhancing educational dynamics integrating large language models with a social robot. In *Proceedings of the 2024 8th International Conference on Software and e-Business*, pages 87–94. ACM.

Tennent, H., Shen, S., and Jung, M. (2019). Micbot: A peripheral robotic object to shape conversational dynamics and team performance. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 133–142.

Traeger, M. L., Sebo, S. S., Jung, M., Scassellati, B., and Christakis, N. A. (2020). Vulnerable robots positively shape human conversational dynamics in a human-robot team. *Proceedings of the National Academy of Sciences*, 117(12):6370–6375.

Weisswange, T. H., Javed, H., Dietrich, M., Jung, M. F., and Jamali, N. (2026). Design implications for robots that facilitate groups—a scoping review on improving group interactions through directed robot action. *J. Hum.-Robot Interact.*, 15(2).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, 3:160018.

A Consent Form

INFORMED CONSENT FORM

You are being invited to participate in a research study about creative collaboration with one or more social robots. This study is being done by Ruben Weijers and Catharine Oertel from the TU Delft.

The purpose of this study is to understand the effect of a social robot's interaction style on human collaboration. The session will take approximately 30-50 minutes. The data will be used for BSc theses and potential publication. You will be asked to complete a brief questionnaire, collaborate with a human partner and a social robot on an open-ended challenge, complete further questionnaires, and take part in a short group interview with your partner about your experience. During the session, we will collect: (1) audio and/or video recordings of the session, (2) your responses to questionnaires, and (4) basic demographic information (such as age, gender, and country of origin) used only to describe the overall participant sample.

To the best of our ability, your answers in this study will remain confidential. We will minimize any risk by removing any mention of names or sensitive information from data.

Your participation is entirely voluntary and you may withdraw at any time during the session without giving any reason. During the session, you are free to stop at any time without providing a reason, and you are free to request the deletion of your data. You will not be financially compensated for your time.

	Yes	No
A: GENERAL AGREEMENT – RESEARCH GOALS, PARTICIPANT TASKS AND VOLUNTARY PARTICIPATION		
I have read and understood the above information.	☐	☐

For questions or requests to delete your data, contact: r.weijers@tudelft.nl

PLEASE TICK THE APPROPRIATE BOXES:

I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
I consent voluntarily to be a participant in this study and understand that I can withdraw from the study at any time, without having to give a reason.	☐	☐
I understand that taking part in the study involves discussion with a conversational robot.	☐	☐
I understand that taking part in the study involves completing questionnaires and a short group interview about my experience.	☐	☐

I understand that the interview takes place with my partner present, and that I should not share anything I would not want my partner to hear.	☐	☐
I understand that the study will last approximately 45 minutes.	☐	☐
I understand that the session will be audio and video recorded	☐	☐
B: POTENTIAL RISKS OF PARTICIPATING (INCLUDING DATA PROTECTION)		
I understand that my data will be treated confidentially, that any direct identifiers (such as my name) will be replaced by a pseudonym for analysis, and that names mentioned during the session will be removed from transcripts.	☐	☐

I understand that I may request deletion of my data up until June 15th, after which deletion may no longer be possible	☐	☐
I understand that I must not provide any personally identifiable information such as phone number, email address or password. If I do this, it will be removed from the recordings and this may destroy the consistency of the data.	☐	☐
I understand that anonymised research data will be stored for 10 years in accordance with TU Delft's Research Data Framework Policy.	☐	☐
I understand that anonymised outputs from this study, including redacted transcripts, coded behavioural data, aggregated survey responses, and individually screened quotes, may be shared with other researchers on request, under a Creative Commons Attribution (CC BY 4.0) licence requiring attribution to the original researchers. I understand that raw audio and video recordings will not be shared outside the research team.	☐	☐
C: RESEARCH PUBLICATION, DISSEMINATION AND APPLICATION		
I understand that after the research study the de-identified information I provide will be used for BSc theses / potential publications.	☐	☐
I agree that my responses can be quoted anonymously in research outputs.	☐	☐
I agree that some parts of the conversation and task outcome can be shown in research outputs (BSc theses, potential publications) or snapshots of them can appear anonymously.	☐	☐

D: (LONGTERM) DATA STORAGE, ACCESS AND REUSE		
I give permission for the anonymised data that I provide to be archived in the 4TU repository so it can be used for future research and learning.	☐	☐
I understand that access to this repository is restricted and that other researchers may request access for non-commercial research and teaching purposes.	☐	☐

Signatures

____________________________ ____________________ _________

Name of participant Signature Date

I, as researcher, have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

__________________________ ___________________ _________

Researcher name Signature Date

Study contact details for further information: Ruben Weijers, r.weijers@tudelft.nl

B Task Instructions

Task Description

Your group task is to brainstorm ideas for:

"Improving Campus Life"

Try to think broadly about ways to improve the experience of students on campus.

Possible topics:

- student well-being
- social interaction
- food and services
- events and activities
- safety and comfort

There are no correct or incorrect answers. The goal is to generate as many meaningful and creative ideas as possible.

Interaction With Pepper

Pepper is able to participate in the discussion by providing solutions.

You may interact with Pepper naturally during the session. If you would like Pepper to contribute, you can address him directly by saying "Pepper".

Please continue the discussion as naturally as possible, regardless of whether Pepper speaks.

Instructions During the Task

Please:

- discuss ideas openly
- try to build upon each other's suggestions
- talk loud and clear

Please avoid:

- intentionally testing or distracting the robot
- speaking over other participants

Duration

The session will last approximately 12 minutes.

After the brainstorming session, you will be asked to complete two short questionnaires regarding your experience during the interaction and take part in an interview about how you perceived your partner and the robot.

Questions

If you have any questions before or after the session, please ask us.

Thank you for participating.

C Post-Session Questionnaire

Post-Session Questionnaire

Thank you for participating. This short questionnaire asks about your experience during the session. There are no right or wrong answers - please answer based on your honest impression. Your responses are confidential and will only be used for research purposes. It takes approximately 5 minutes to complete.

* Required

What is your age? *

Enter your answer

What is your gender? *

- ○ Woman
- ○ Man
- ○ Non-binary
- ○ Prefer not to say

What is the highest level of education you are currently pursuing or have completed? *

- ○ PhD
- ○ Master
- ○ Bachelor
- ○ Other

How well did you know your partner before this session? *(1 = Never met before → 5 = Know them very well)* *

1	2	3	4	5
○	○	○	○	○

How many times have you worked on a task together before? *(1 = Never, 2= Once or twice, 3 = Several times, 4 = Regularly)* *

1	2	3	4
○	○	○	○

*

	1 = *Strongly Disagree*	2	3	4	5	6	7 = *Strongly Agree*
I see myself as extraverted, enthusiastic.	○	○	○	○	○	○	○
I see myself as reserved, quiet.	○	○	○	○	○	○	○

How much experience do you have with conversational AI tools (e.g. ChatGPT, Siri, Alexa)? *

Never	Once or twice	Several times	Regularly
○	○	○	○

Your Experience as a Group

The following statements are about how you felt as a group **during this session**. Please indicate how much you agree with each statement. *(1 = Strongly Disagree, 7 = Strongly Agree)* *

	1	2	3	4	5	6	7
1. I feel that I belong to this group.	○	○	○	○	○	○	○
2. I am happy to be part of this group.	○	○	○	○	○	○	○
3. I see myself as part of this group.	○	○	○	○	○	○	○
4. This group is one of the best anywhere.	○	○	○	○	○	○	○
5. I feel that I am a member of this group.	○	○	○	○	○	○	○
6. I am content to be part of this group.	○	○	○	○	○	○	○

Your Sense of Group Membership

The following statements are about how you identified with the group you were just part of. Please respond based on your experience **during this session**. *(1 = Strongly Disagree, 7 = Strongly Agree)* *

	1	2	3	4	5	6	7
1. I would prefer to be in a different group	○	○	○	○	○	○	○
2. In this group, members don't have to rely on one another	○	○	○	○	○	○	○
3. I think of this group as part of who I am	○	○	○	○	○	○	○
4. Members of this group like one another	○	○	○	○	○	○	○
5. All members need to contribute to achieve the group's goals	○	○	○	○	○	○	○
6. I see myself as quite different from other members of the group	○	○	○	○	○	○	○
7. I enjoy interacting with the members of this group	○	○	○	○	○	○	○
8. This group accomplishes things that no single member could achieve	○	○	○	○	○	○	○
9. I don't think of this group as part of who I am	○	○	○	○	○	○	○
10. I don't like many of the other people in this group	○	○	○	○	○	○	○
11. In this group, members do not need to cooperate to complete group tasks	○	○	○	○	○	○	○
12. I see myself as quite similar to other members of the group	○	○	○	○	○	○	○

D Interview

Before you start:

"Thank you for taking part. I want to ask you a few questions about your experience in the session - there are no right or wrong answers, I'm really just curious about what it felt like from the inside. I'll video and audio record the interview if that's alright. This will take about five to ten minutes."

Block 1 - General warm-up

Q1. "How did the conversation go for you? What stood out?"

- If they only mention the robot → note it, but don't probe yet, let them talk.
- If they mention their partner unprompted → follow immediately: *"Tell me more about that."*

Follow-up A: "Was there a moment that felt particularly natural, or one that felt a bit off?"

- If they describe a specific moment, ask: *"What was happening just before that?"*

Follow-up B: "Did anything about the session surprise you compared to what you expected?"

- Listen for surprises about the robot, about their partner, or about themselves. I will code these separately in my analysis.

Block 2 - Robot perception

Q2. "How would you describe what the robot was doing in the conversation? What role did it seem to have?"

Follow-up A: "Did it feel more like a participant in the conversation or more like something directing it from the outside?"

- If they hesitate: *"Like, did it feel like a third person talking, or more like a moderator?"*

Follow-up B: "Were there moments when the robot said something unexpected?"

- If yes: *"What did you do after that?"* - searching for behavioural response, not just evaluation.

Follow-up C: "Did you find yourself adjusting how you were speaking because of something the robot did?"

- If yes: *"Can you give me an example?"*
- If no: file that as data too, especially if the transcript shows otherwise.

Block 3 - Partner perception

Q3. "Thinking about your partner - how would you describe how the two of you talked to each other during the session?"

Let them answer fully before any follow-up. First unprompted words here are the richest data.

Follow-up A: "Did you feel like you were both contributing roughly equally, or did one of you take more of a lead?"

- If unequal: *"Did that feel natural, or a bit uncomfortable?"*
- If equal: *"Was there a point where that shifted?"*

Follow-up B: "Did you feel like you were building on each other's ideas, or were you more taking turns sharing separate things?"

- If building on: *"Can you give me an example of a moment like that?"*
- If separate: *"Do you think the topic had anything to do with that, or was it more just how the conversation flowed?"*

Follow-up C: "Were there moments where you felt genuinely connected to what your partner was saying - like it really landed with you?"

- If yes: *"What was that about?"*
- If they describe a campus-specific moment (housing, stress, social life): note the topic because it may cluster thematically across conditions.

Q4. "Did anything about how you and your partner talked feel different from how you'd normally chat with someone about this kind of topic?"

Do not rush to a follow-up.

Follow-up A: "Can you think of a specific moment that felt that way?"

- If yes: *"What do you think caused it?"* Wait for them to name a cause before suggesting anything.

Follow-up B: "Was it more something your partner did, something the robot did, or just how the situation unfolded?"

- Do not offer these options until they are stuck. If they say "I'm not sure," then offer the three options as a prompt. Unprompted attribution is more valuable.

Follow-up C: "Did you find yourself more focused on your partner during this conversation, or more focused on the robot?"

- If more on robot: *"Did that change how much you paid attention to what your partner was saying?"*
- If more on partner: *"Did the robot feel like it was getting in the way at any point, or did it not really affect that?"*

Follow-up D: "Campus experience is something most people have pretty personal views about. Did having the robot there change how open you felt sharing your own experiences compared to if it were just you and your partner?"

Block 4 - Counterfactual

Q5. "If you had had the same conversation but without the robot (just you and your partner) do you think it would have gone the same way?"

Follow-up A: "What specifically do you think would have been different?"

- Push for concrete examples: *"Like in terms of who talked more, or what you talked about, or how you responded to each other?"*

Follow-up B: "Do you think you would have gotten into the same depth on any of the topics, or would it have stayed more surface level?"

Follow-up C: "Do you think the way you felt about your partner after the session would have been different without the robot?"

Block 5 - Closing

Q6. "Is there anything about the conversation (about the robot, about your partner, about the campus topics you got into) that struck you but we haven't talked about?"

- If they bring up something new here: *"Tell me more about that."* Don't do anything else.

E Coding Manual

Mercer Talk Types Coding Manual

Purpose

This manual provides simple coding rules for classifying conversation episodes as Disputational, Cumulative, or Exploratory Talk based on Mercer's framework. Code short discussion episodes (2 turns) rather than individual utterances.

1. Unit of Analysis

- Code an episode, not a single sentence.
- An episode is a discussion around one idea, decision, disagreement, or problem.
- If multiple talk types appear, code the dominant type.
- If uncertain, prioritize: Exploratory > Cumulative > Disputational.

2. Quick Decision Tree

A. Do participants explain WHY, justify ideas, or evaluate alternatives?
→ Exploratory

B. If not, are participants mainly disagreeing or rejecting ideas?
→ Disputational

C. If not, are participants agreeing and building on ideas?
→ Cumulative

3. Disputational Talk

Definition: Disagreement without constructive reasoning.

Indicators:
- Rejections
- Contradictions
- Competition
- Assertions without explanation

Typical phrases:
"No." "That's wrong." "Do mine."

Example 1

A: We should make the robot ask questions.
B: No, it shouldn't.
A: Yes it should.
B: No.

Decision: DISPUTATIONAL
Reason: Repeated disagreement with no justification.

4. Cumulative Talk

Definition: Participants accept and build on ideas without critical evaluation.

Indicators:
- Agreement
- Confirmation
- Repetition

- Elaboration

Typical phrases:
"Yeah." "Exactly." "Good idea."

Example 2

A: The robot could remind people to speak.
B: Yeah.
A: And it could ask quieter people for ideas.
B: Yeah, that's good.
A: That would help everyone participate.
B: Exactly.

Decision: CUMULATIVE
Reason: Ideas are accepted and expanded, but not evaluated.

5. Exploratory Talk

Definition: Participants critically but constructively examine ideas and provide reasons.

Indicators:
- Explanations
- Justifications
- Alternative suggestions
- Evaluation of ideas
- Joint decision making

Typical phrases:
"Because..." "I think..." "What if..."

Example 3

A: The robot should ask quieter people for ideas.
B: Maybe, but it could make them uncomfortable.
A: That's true. What if it asked everyone in turn instead?
B: That might work better because no one feels singled out.
A: Yes, and it would be fairer.

Decision: EXPLORATORY
Reason: Participants challenge, justify, evaluate, and refine ideas together.

6. Reliability Rules

1. Look for explicit reasoning first.
2. If reasoning is present, prefer Exploratory.
3. If disagreement dominates and no reasons are given, use Disputational.
4. If participants mainly agree and build on ideas, use Cumulative.
5. Record a short justification for every coding decision.

F Justifications for Usage of Covariates

This appendix reports the preliminary linear regression models used to determine which covariates warranted inclusion in the primary analyses.

F.1 Justification of Usage of Covariates in Participation Balance

Tables 3 and 4 report the regression model predicting adjusted Gini index from strategy, familiarity, extraversion, and prior AI experience. No predictor reached significance, so all covariates were excluded from the primary participation balance analysis.

Table 3: Linear regression model fit statistics for adjusted Gini index used in covariate screening.

Model	N	R	R^2
1.00	20	0.35	0.12

Table 4: Linear regression coefficients for the model predicting adjusted Gini index from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	0.30	0.43	0.69	.502
Familiarity	0.04	0.08	0.51	.620
Extravertness	0.02	0.05	0.48	.640
AI Experience	-0.09	0.09	-1.05	.311
Strategy:				
Assertive – Supportive	0.07	0.10	0.70	.496

F.2 Justification of Usage of Covariates in Collaborative Talk Types

Tables 5 through 10 report the regression models predicting each Mercer talk type. No covariate reached significance for any talk type, so covariates were excluded from the primary one-way ANOVA tests.

Table 5: Linear regression model fit statistics for disputational talk type used in covariate screening.

Model	N	R	R^2
1.00	20	0.51	0.26

Table 6: Linear regression coefficients for the model predicting disputational talk type from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	0.05	0.06	0.75	.462
Familiarity	-0.02	0.01	-1.37	.191
Extravertness	0.01	0.01	1.30	.214
AI Experience	-3.30e-4	0.01	-0.03	.979
Strategy:				
Assertive – Supportive	0.01	0.01	0.92	.372

Table 7: Linear regression model fit statistics for cumulative talk type used in covariate screening.

Model	N	R	R^2
1.00	20	0.43	0.18

Table 8: Linear regression coefficients for the model predicting cumulative talk type from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	0.02	0.08	0.21	.838
Familiarity	0.01	0.01	0.46	.651
Extravertness	-0.01	0.01	-1.12	.279
AI Experience	0.03	0.02	1.57	.137
Strategy:				
Assertive – Supportive	-0.01	0.02	-0.42	.684

Table 9: Linear regression model fit statistics for exploratory talk type used in covariate screening.

Model	N	R	R^2
1.00	20	0.65	0.42

Table 10: Linear regression coefficients for the model predicting exploratory talk type from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	0.35	0.14	2.49	.025
Familiarity	-0.03	0.03	-1.24	.234
Extravertness	0.03	0.02	1.68	.113
AI Experience	-0.02	0.03	-0.61	.548
Strategy:				
Assertive – Supportive	-0.08	0.03	-2.52	.024

F.3 Justification of Usage of Covariates in Cohesion Scale

Tables 11 through 14 report the regression models predicting the Belonging and Morale sub-scales. AI experience reached significance for Belonging ($p = .046$). All covariates were retained in that ANCOVA model.

Table 11: Linear regression model fit statistics for Belonging sub-scale used in covariate screening.

Model	N	R	R^2
1.00	40	0.37	0.14

Table 12: Linear regression coefficients for the model predicting Belonging sub-scale from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	2.27	1.50	1.51	.140
Familiarity	0.28	0.30	0.94	.356
Extravertness	0.10	0.15	0.68	.499
AI Experience	0.49	0.24	2.06	.046
Strategy:				
Assertive – Supportive	-0.22	0.39	-0.57	.575

Table 13: Linear regression model fit statistics for Morale sub-scale used in covariate screening.

Model	N	R	R^2
1.00	40	0.32	0.10

Table 14: Linear regression coefficients for the model predicting Morale sub-scale from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	2.37	1.65	1.44	.160
Familiarity	0.38	0.33	1.15	.257
Extravertness	-0.10	0.16	-0.61	.547
AI Experience	0.38	0.26	1.48	.149
Strategy:				
Assertive – Supportive	0.24	0.43	0.55	.583

F.4 Justification of Usage of Covariates in Group Identification Scale

Tables 15 through 20 report the regression models predicting the Affective, Behavioral, and Cognitive sub-scales. Familiarity reached significance for the Cognitive sub-scale (p = .013). All covariates were retained in that model.

Table 15: Linear regression model fit statistics for Affective sub-scale used in covariate screening.

Model	N	R	R^2
1.00	40	0.21	0.05

Table 16: Linear regression coefficients for the model predicting Affective sub-scale from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	4.14	1.40	2.96	.005
Familiarity	0.16	0.28	0.56	.577
Extravertness	0.01	0.14	0.08	.937
AI Experience	0.22	0.22	0.98	.334
Strategy:				
Assertive – Supportive	0.14	0.36	0.38	.706

Table 17: Linear regression model fit statistics for Behavioral sub-scale used in covariate screening.

Model	N	R	R^2
1.00	40	0.24	0.06

Table 18: Linear regression coefficients for the model predicting Behavioral sub-scale from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	5.77	1.27	4.56	$< .001$
Familiarity	-0.28	0.25	-1.11	.275
Extravertness	0.05	0.12	0.44	.661
AI Experience	-0.06	0.20	-0.29	.771
Strategy:				
Assertive – Supportive	-0.23	0.33	-0.70	.490

Table 19: Linear regression model fit statistics for Cognitive sub-scale used in covariate screening.

Model	N	R	R^2
1.00	40	0.41	0.17

Table 20: Linear regression coefficients for the model predicting Cognitive sub-scale from strategy, familiarity, extraversion, and prior AI experience.

Predictor	Estimate	SE	t	p
Intercept	7.14	1.43	4.98	$< .001$
Familiarity	-0.74	0.28	-2.61	.013
Extravertness	-0.03	0.14	-0.22	.830
AI Experience	-0.13	0.23	-0.56	.579
Strategy:				
Assertive – Supportive	0.38	0.37	1.02	.316

G Participation Balance Other Tests

This appendix reports the assumption checks for the participation balance analysis. Levene's test confirmed equal variances across conditions. The Shapiro-Wilk test indicated a violation of normality ($p = .026$), which motivated the use of a Mann-Whitney U test in place of a standard ANOVA. The Q-Q plot in Figure 6 illustrates this violation.

Table 21: Levene's test assessing the homogeneity of variances assumption for participation balance (adjusted Gini index) between robot interaction strategy conditions.

	F	df1	df2	p
Strategy	0.07	1	18	.791

Table 22: Shapiro-Wilk test assessing the normality assumption for participation balance scores measured using the adjusted Gini index. The test shows that the normality assumption is violated.

	W	p
Strategy	0.89	.026

Note. A low p-value suggests a violation of the assumption of normality

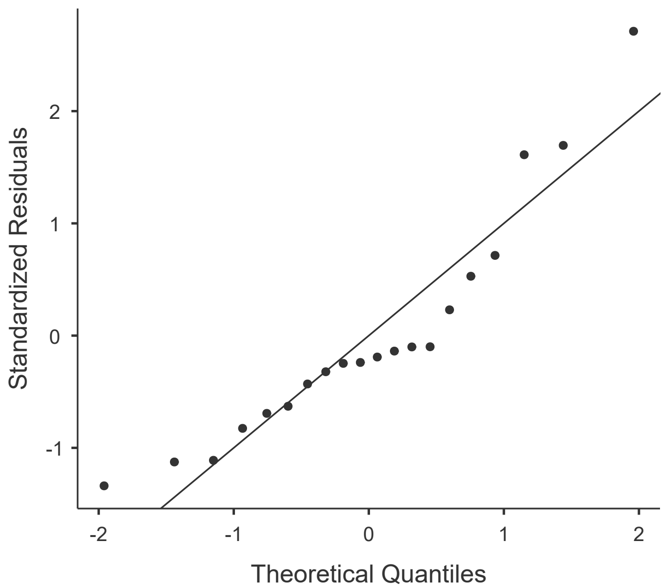


Figure 6: Q-Q plot of the residuals of the Adjusted Gini Index. The plot shows the normality assumption violation of the residuals.

H Talk Types

This appendix reports the assumption checks for the Mercer talk type analyses. Shapiro-Wilk tests confirmed normality for all three talk types, and Levene's tests confirmed homogeneity of variances. These results justified the use of one-way ANOVAs as the primary tests. The Q-Q plots in Figure 7 support these conclusions visually.

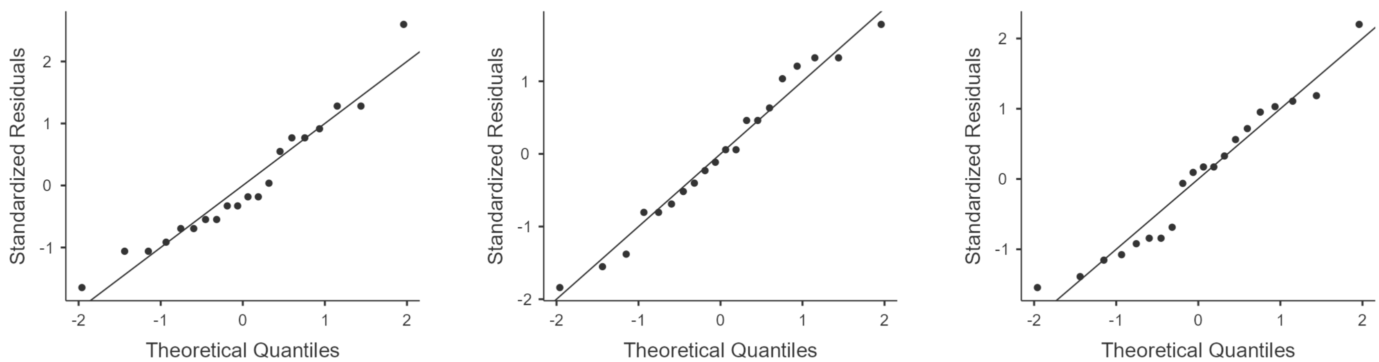
Table 23: Results of Shapiro-Wilk tests assessing the normality assumption for the proportions of disputational, cumulative, and exploratory talk across dyads. No significant deviations from normality were detected.

	W	p
Disputational	0.94	.289
Cumulative	0.97	.821
Exploratory	0.95	.437

Note. A low p-value suggests a violation of the assumption of normality

Table 24: Results of Levene's tests assessing the homogeneity of variance assumption for the proportions of disputational, cumulative, and exploratory talk across interaction strategy conditions. The assumption of equal variances was met for all talk types.

	F	df1	df2	p
Disputational	0.01	1	18	.915
Cumulative	0.45	1	18	.511
Exploratory	1.27	1	18	.275



(a) Disputational Talk (b) Cumulative Talk (c) Exploratory Talk

Figure 7: Normal Q-Q plots of residuals for discourse metrics. The alignment of points along the diagonal indicates that the residuals are normally distributed.

I Cohesion Scale Results

Tables 25 and 26 report the full ANCOVA results for the Belonging and Morale sub-scales of the Perceived Cohesion Scale. Strategy did not significantly affect either sub-scale. AI experience reached significance for Belonging, indicating that participants with more conversational AI experience reported stronger group belonging regardless of condition.

Table 25: ANCOVA results for the Belonging sub-scale, with interaction strategy as the independent variable and familiarity, extravertness, and AI experience as covariates.

	F (1, 35)	p	η^2p	ω^2
Strategy	0.32	.575	0.009	-0.016
Familiarity	0.87	.356	0.024	-0.003
Extravertness	0.47	.499	0.013	-0.013
AI Experience	4.26	.046	0.109	0.078

Table 26: ANCOVA results for the Morale sub-scale, with interaction strategy as the independent variable and familiarity, extravertness, and AI experience as covariates.

	F (1, 35)	p	η^2p	ω^2
Strategy	0.31	.583	0.009	-0.017
Familiarity	1.33	.257	0.037	0.008
Extravertness	0.37	.547	0.010	-0.016
AI Experience	2.18	.149	0.059	0.029

J Cohesion Scale Reliability Results

This appendix reports internal consistency statistics for the Perceived Cohesion Scale and its sub-scales. The overall scale and both sub-scales showed strong reliability, with Cronbach's α and McDonald's ω both above .80 across all components.

Table 27: Internal consistency reliability statistics for the Cohesion scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Cohesion Scale	0.90	0.91

Table 28: Item-level reliability statistics for the Cohesion scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Belonging 1	0.71	0.89	0.89
Morale 1	0.75	0.88	0.89
Belonging 2	0.79	0.88	0.88
Morale 2	0.65	0.90	0.90
Belonging 3	0.72	0.88	0.89
Morale 3	0.79	0.87	0.88

Table 29: Internal consistency reliability statistics for the Belonging scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Belonging Scale	0.83	0.84

Table 30: Item-level reliability statistics for the Belonging scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Belonging 1	0.65	0.81	0.81
Belonging 2	0.79	0.67	0.67
Belonging 3	0.64	0.81	0.82

Table 31: Internal consistency reliability statistics for the Morale scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Morale Scale	0.81	0.83

Table 32: Item-level reliability statistics for the Morale scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Morale 1	0.67	0.74	0.75
Morale 2	0.60	0.82	0.82
Morale 3	0.73	0.66	0.68

K Group Identification Scale Results

Tables 33 through 35 report the full ANCOVA results for the Affective, Behavioral, and Cognitive sub-scales of the Tripartite Model scale. Strategy did not significantly affect any sub-scale. Familiarity reached significance for the Cognitive sub-scale, indicating that participants who knew each other beforehand reported stronger cognitive identification with the group.

Table 33: ANCOVA results for the Affective sub-scale, with interaction strategy as the independent variable and familiarity, extravertness, and AI experience as covariates.

	F (1, 35)	p	η^2p	ω^2
Strategy	0.14	.706	0.004	-0.023
Familiarity	0.32	.577	0.009	-0.018
Extravertness	0.01	.937	0.000	-0.027
AI Experience	0.96	.334	0.027	-0.001

Table 34: ANCOVA results for the Behavioral sub-scale, with interaction strategy as the independent variable and familiarity, extravertness, and AI experience as covariates.

	F (1, 35)	p	η^2p	ω^2
Strategy	0.49	.490	0.014	-0.013
Familiarity	1.23	.275	0.034	0.006
Extravertness	0.20	.661	0.006	-0.021
AI Experience	0.09	.771	0.002	-0.024

Table 35: ANCOVA results for the Cognitive sub-scale, with interaction strategy as the independent variable and familiarity, extravertness, and AI experience as covariates.

	F (1, 35)	p	η^2p	ω^2
Strategy	1.03	.316	0.029	0.001
Familiarity	6.80	.013	0.163	0.131
Extravertness	0.05	.830	0.001	-0.022
AI Experience	0.31	.579	0.009	-0.016

L Group Identification Scale Reliability Results

This appendix reports internal consistency statistics for the Group Identification scale and its sub-scales. The overall scale showed acceptable reliability. The Behavioral sub-scale showed poor internal consistency (Cronbach's α = .41, McDonald's ω = .53) and its results should be interpreted with caution.

Table 36: Internal consistency reliability statistics for the Group Identification scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Group Identification Scale	0.71	0.75

Table 37: Item-level reliability statistics for the Group Identification scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Affective 1^R	0.53	0.66	0.71
Behavioral 1^R	-0.11	0.76	0.78
Cognitive 1	0.42	0.68	0.73
Affective 2	0.54	0.67	0.70
Behavioral 2	0.50	0.67	0.71
Cognitive 2^R	0.00	0.74	0.77
Affective 3	0.56	0.67	0.70
Behavioral 3	0.29	0.70	0.74
Cognitive 3^R	0.53	0.66	0.72
Affective 4^R	0.34	0.69	0.74
Behavioral 4^R	0.24	0.71	0.75
Cognitive 4	0.47	0.68	0.72

^{R} reverse scaled item

Table 38: Internal consistency reliability statistics for the Affective scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Affective Scale	0.75	0.78

Table 39: Item-level reliability statistics for the Affective scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Affective 1^R	0.49	0.73	0.76
Affective 2	0.64	0.64	0.71
Affective 3	0.75	0.58	0.66
Affective 4^R	0.34	0.79	0.81

^{R} reverse scaled item

Table 40: Internal consistency reliability statistics for the Behavioral scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Behavioral Scale	0.41	0.53

Table 41: Item-level reliability statistics for the Behavioral scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Behavioral 1^R	0.02	0.55	0.61
Behavioral 2	0.30	0.26	0.32
Behavioral 3	0.32	0.24	0.47
Behavioral 4^R	0.31	0.26	0.50

^{R} reverse scaled item

Table 42: Internal consistency reliability statistics for the Cognitive scale, including Cronbach's α and McDonald's ω.

	Cronbach's α	McDonald's ω
Cognitive Scale	0.63	0.69

Table 43: Item-level reliability statistics for the Cognitive scale.

		If item dropped	
	Item-rest correlation	Cronbach's α	McDonald's ω
Cognitive 1	0.60	0.41	0.46
Cognitive 2^R	0.18	0.71	0.76
Cognitive 3^R	0.55	0.45	0.59
Cognitive 4	0.36	0.60	0.72

^{R} reverse scaled item